

# Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/EP04/053471

International filing date: 14 December 2004 (14.12.2004)

Document type: Certified copy of priority document

Document details: Country/Office: EP  
Number: 03293216.2  
Filing date: 18 December 2003 (18.12.2003)

Date of receipt at the International Bureau: 01 April 2005 (01.04.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)



World Intellectual Property Organization (WIPO) - Geneva, Switzerland  
Organisation Mondiale de la Propriété Intellectuelle (OMPI) - Genève, Suisse



**Europäisches  
Patentamt**

**European  
Patent Office**

**Office européen  
des brevets**

**Bescheinigung**

**Certificate**

**Attestation**

Die angehefteten Unterla-  
gen stimmen mit der  
ursprünglich eingereichten  
Fassung der auf dem näch-  
sten Blatt bezeichneten  
europäischen Patentanmel-  
dung überein.

The attached documents  
are exact copies of the  
European patent application  
described on the following  
page, as originally filed.

Les documents fixés à  
cette attestation sont  
conformes à la version  
initialement déposée de  
la demande de brevet  
européen spécifiée à la  
page suivante.

**Patentanmeldung Nr.    Patent application No.    Demande de brevet n°**

03293216.2

Der Präsident des Europäischen Patentamts;  
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets  
p.o.

**R C van Dijk**





Anmeldung Nr:  
Application no.: 03293216.2  
Demande no:

Anmeldetag:  
Date of filing: 18.12.03  
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

Thomson Licensing S.A.  
46, quai A. Le Gallo  
92100 Boulogne-Billancourt  
FRANCE

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:  
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.  
If no title is shown please refer to the description.  
Si aucun titre n'est indiqué se référer à la description.)

Device and method for creating a saliency map of an image

In Anspruch genommene Priorität(en) / Priority(ies) claimed / Priorité(s)  
revendiquée(s)  
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/  
Classification internationale des brevets:

G06T/

Am Anmeldetag benannte Vertragsstaaten/Contracting states designated at date of  
filing/Etats contractants désignées lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IT LU MC NL  
PT RO SE SI SK TR LI



The invention is related to a device and a method for creating a saliency map of an image.

5       The human information processing system is intrinsically a limited system and especially for the visual system. In spite of the limits of our cognitive resources, this system has to face up to a huge amount of information contained in our visual environment. Nevertheless and paradoxically, humans seem to succeed in solving this problem since we are  
10       able to understand our visual environment.

It is commonly assumed that certain visual features are so elementary to the visual system that they require no attentional resources to be perceived. These visual features are called pre-attentive features.

According to this tenet of vision research, human attentive behavior is  
15       shared between pre-attentive and attentive processing. As explained before, pre-attentive processing, so-called bottom-up processing, is linked to involuntary attention. Our attention is effortless drawn to salient parts of our view. When considering attentive processing, so-called top-down processing, it is proved that our attention is linked to a particular task that we have in  
20       mind. This second form of attention is thus a more deliberate and powerful one in the way that this form of attention requires effort to direct our gaze towards a particular direction.

25       The detection of the salient points in an image enable the improvement of further steps such as coding or image indexing, watermarking, video quality estimation.

30       The known approaches are more or less based on non-psycho visual features. In opposition with such methods, the proposed method relies on the fact that the model is fully based on the human visual system (HVS) such as the computation of early visual features.

In a first aspect, the invention proposes a method for creating a saliency map of an image comprising the steps of :

- projection of said image according to the luminance component and if said image is a color image, according to the luminance component and according to the chrominance components,
- perceptual sub-bands decomposition of said components according to the visibility threshold of a human eye,
- extraction of the salient elements of the sub-bands related to the luminance component,
- contour enhancement of said salient elements in each sub-band related to the luminance component,
- calculation of a saliency map from the contour enhancement, for each sub-band related to the luminance component.
- creation of the saliency map as a function of the saliency maps obtained for each sub-band.

In a second aspect, the invention proposes a device for creating a saliency map of an image characterized in that it comprises means for:

- Projecting said image according to the luminance component and if said image is a color image, according to the luminance component and according to the chrominance components,
- Transposing into the frequential domains said luminance and chrominance signals,
- Decomposing into perceptual sub-bands said components of the frequential domain according to the visibility threshold of a human eye,
- Extracting the salient elements of the sub-bands related to the luminance component,
- Contour enhancing said salient elements in each sub-band related to the luminance component,
- Calculating a saliency map from the contour enhancement, for each sub-band related to the luminance component.
- Creating the saliency map as a function of the saliency maps obtained for each sub-band.

Other characteristics and advantages of the invention will appear through the description of a non-limiting embodiment of the invention, which will be illustrated, with the help of the enclosed drawings, wherein:

- 5           -       Figure 1 represents a general flow-chart of a preferred embodiment of the method according to the invention applied to a black and white image,
- Figure 2 represents a general flow-chart of a preferred embodiment of the method according to the invention applied to a black and white image,
- 10          -       Figure 3 represents the psycho visual spatial frequency partitioning for the achromatic component,
- Figure 4 represents the psycho visual spatial frequency partitioning for the chromatic components,
- 15          -       Figure 5 represents the Dally Contrast Sensitivity Function,
- Figure 6a and 6b represent respectively the visual masking and a non linear model of masking,
- Figure 7 represents the flow-chart of the normalisation step according to the preferred embodiment,
- 20          -       Figure 8 represents the inhibition/excitation step,
- Figure 9 represents the profile of the filters to model facilitative interactions for  $\theta=0$ ,
- Figure 10 represents an illustration of the operator  $D(z)$ ,
- Figure 11 represents the chromatic reinforcement step,
- 25          -       Figure 12 represents the non CRF exhibition caused by the adjacent areas of the CRF flanks,
- Figure 13 represents a profile example of the normalized weighting function for a particular orientation and radial frequency.

30

Figure 1 represents the general flow-chart of a preferred embodiment of the method according to the invention applied to a black and white image.

The algorithm is divided in three main parts.



The first one named visibility is based on the fact that the human visual system (HVS) has a limited sensitivity. For example, the HVS is not able to perceive with a good precision all signals in your real environment and is insensible to small stimuli. The goal of this first step has to take into account these intrinsic limitations by using perceptual decomposition, contrast sensitivity functions (CSF) and masking functions.

The second part is dedicated to the perception concept. The perception is a process that produces from images of the external world a description that is useful to the viewer and not cluttered with irrelevant information. To select relevant information, a center surround mechanism is notably used in accordance with biological evidences.

The last step concerns some aspects of the perceptual grouping domain. The perceptual grouping refers to the human visual ability to extract significant images relations from lower level primitive image features without any knowledge of the image content and group them to obtain meaningful higher-level structure. The proposed method just focuses on contour integration and edge linking.

Steps E3, E4 are executed on the signal in the frequential domain.

Steps E1, E6 and E9 are done in the spatial domain.

Steps E7 and E8 are done in the frequential or spatial domain. If they are done in the frequential domain, a Fourier transformation has to be carried on before step E7 and an inverse Fourier transformation has to be carried out before step E9.

In step E1, the luminance component is extracted from the considered image.

In step E2, the luminance component is transposed into the frequency domain by using known transformations such as the Fourier transformation in order to be able to apply in step E3, the perceptual sub-band decomposition on the image.

In step E3, a perceptual decomposition is applied on the luminance component. This decomposition is inspired from the cortex transform and based on the decomposition proposed in the document "The computation of visual bandwidths and their impact in image decomposition and coding", International Conference and Signal Processing Applications and Technology,

Santa-Clara, California, pp. 776-770, 1993. This decomposition is done according to the visibility threshold of a human eye.

5 The decomposition, based on different psychophysics experiments, is obtained by carving up the frequency domain both in spatial radial frequency and orientation. The perceptual decomposition of the component A leads to 17 psycho visual sub-bands distributed on 4 crowns as shown on figure 3.

The shaded region on the figure 3 indicates the spectral support of the sub-band belonging to the third crown and having an angular selectivity of 30 degrees, from 15 to 45 degrees.

10 Four domains (crowns) of spatial frequency are labeled from I to IV:

I: spatial frequencies from 0 to 1.5 cycles per degree;

II: spatial frequencies from 1.5 to 5.7 cycles per degree;

III: spatial frequencies from 5.7 to 14.2 cycles per degree;

IV: spatial frequencies from 14.2 to 28.2 cycles per degree.

15

The angular selectivity depends on the considered frequency domain. For low frequencies, there is no angular selectivity.

20 The main properties of these decompositions and the main differences from the cortex transform are a non-dyadic radial selectivity and an orientation selectivity that increases with the radial frequency.

25 Each resulting sub-band may be regarded as the neural image corresponding to a population of visual cells tuned to a range of spatial frequency and a particular orientation. In fact, those cells belong to the primary visual cortex (also called striate cortex or V1 for visual area 1). It consists of about 200 million neurons in total and receives its input from the lateral geniculate nucleus. About 80 percent of the cells are selective for orientation and spatial frequency of the visual stimulus.

30 On the image spatial spectrum, a well-known property of the HVS is applied, which is known as the contrast sensitivity function (CSF). The CSF applied is a multivariate function mainly depending on the spatial frequency, the orientation and the viewing distance.

Biological evidences have shown that visual cells response to stimuli above a certain contrast. The contrast value for which a visual cell response is called the visibility threshold (above this threshold, the stimuli is visible). This threshold varies with numerous parameters such as the spatial frequency of the stimuli, the orientation of the stimuli, the viewing distance, ... This variability leads us to the concept of the CSF which expresses the sensitivity of the human eyes (the sensitivity is equal to the inverse of the contrast threshold) as a multivariate function. Consequently, the CSF permits to assess the sensitivity of the human eyes for a given stimuli.

In step E4, a 2D anisotropic CSF designed by Dally is applied. Such a CSF is described in document "the visible different predictor: an algorithm for the assessment of image fidelity", in proceedings of SPIE Human vision, visual processing and digital display III, volume 1666, pages 2-15, 1992.

The CSF enables the modelisation of an important property of the eyes, as the SVH cells are very sensitive to the spatial frequencies.

On figure 5, the Dally CSF is illustrated.

Once the Dally function has been applied, an inverse Fourier Transformation is applied on the signal in step E5, in order to be able to apply the next step E6.

For natural pictures, the sensitivity can be modulated (increased or decreased the visibility threshold) by the presence of another stimulus. This modulation of the sensitivity of the human eyes is called the visual masking, as done in step E6.

An illustration of masking effect is shown on the figures 6a and 6b. Two cues are considered, a target and a masker where  $C_T$  and  $C_M$  are the contrast threshold of the target in the presence of the masker and the contrast of the masker respectively. Moreover,  $C_{T0}$  is the contrast threshold measured by a CSF (without masking effect).

When  $C_M$  varies, three regions can be defined :

- At low values of  $C_M$ , the detection threshold remains constant. The visibility of the target is not modified by the masker.
- When  $C_M$  tends toward  $C_{T0}$ , the masker eases the detection of the target by decreasing the visibility threshold. This phenomenon is called facilitative or pedestal effect.
- When  $C_M$  increases, the target is masked by the masker. His contrast threshold increases.

The visual masking method is based on the detection of a simple signal as sinusoidal patterns.

There are several other methods to achieve the visual masking modeling based on psychophysics experiments: for instance, a best method refers to the detection of quantization noise.

It is obvious that the preferred method is a strong simplification as regard the intrinsic complexity of natural pictures. Nevertheless, numerous applications (watermarking, video quality assessment) are built around such principle with interesting results compared to the complexity.

In the context of sub-band decomposition, masking has been intensively studied leading to define three kinds of masking: intra-channel masking, inter-channel masking and inter-component masking.

The intra-channel masking occurs between signals having the same features (frequency and orientation) and consequently belonging to the same channel. It is the most important masking effect.

The inter-channel masking occurs between signals belonging to different channels of the same component.

The inter-component masking occurs between channels of different components (the component A and one chromatic component for example). These two last visual masking are put together and are just called inter-masking in the following.

For the achromatic component, we used the masking function designed by Dally in document entitled "A visual model for Optimizing the Design of Image Processing Algorithms", in IEEE international conferences on image processing, pages 16-20, 1994, in spite of the fact that this model doesn't take



into account the pedestal effect. The strength of this model lies in the fact that it has been optimized with a huge amount of experimental results.

The variation of the visibility threshold is given by :

$$T_{i,j,A}^{intra}(m,n) = (1 + (k_1(k_2 |R_{i,j}(m,n)|)^s)^b)^{1/b}$$

where  $R_{i,j}$  is a psycho visual channel stemming from the perceptual channel decomposition (For example, the shaded region on the figure 2.1 leads to the channel  $R_{III,2}$ ). The values  $k_1$ ,  $k_2$ ,  $s$ ,  $b$  are given below :

$$k_1 = 0.0153$$

$$k_2 = 392.5$$

The below table gives the values of  $s$  and  $b$  according to the considered sub-band:

Sub-band	$s$	$b$
I	0.75	4
II	1	4
III	0.85	4
IV	0.85	4

We get the signal  $R'_{i,j}(x,y)$  at the output of the masking step.

$$R'_{i,j}(x,y) = R_{i,j}(x,y) / T_{i,j}(x,y)$$

Then, in step E7, the step of normalization enables to extract the main important information from the sub-band. Step E7 is detailed on figure 7.

In step S1, a first sub-band  $R'_{I,j}(x,y)$  is selected. The steps S2 to S4 and S8 are carried on for each sub-band  $R'_{i,j}(x,y)$  of the 17 sub-bands.

The steps S5 to S7 are done for the second crown (II).

I represents the spatial radial frequency band, I belongs to {I, II, III, IV}.

J represents the orientation, j belongs to {1, 2, 3, 4, 5, 6},

(x,y) represent the spatial coordinates.

In other embodiments, the different steps can be carried out on all the sub-bands.

Steps S2 and S3 aim to modelize the behavior of the classical receptive field (CRF).

5 The concept of CRF permits to establish a link between a retinal image and the global percept of the scene. The CRF is defined as a particular region of visual field within which an appropriate stimulation (with preferred orientation and frequency) provokes a relevant response stemming from visual cell. Consequently, by definition, a stimulus in the outer region (called  
10 surround) cannot activate the cell directly.

The inhibition and excitation in steps S2 and S3 are obtained by a Gabor filter, which is sensible as for the orientation and the frequency.

The Gabor filter can be represented as following:

15 
$$gabor(x, y, \sigma_x, \sigma_y, f, \theta) = G_{\sigma_x, \sigma_y}(x_\theta, y_\theta) \cos(2\pi f x_\theta)$$

f being the spatial frequency of the cosinus modulation in cycles per degree (cy/°).

(x<sub>θ</sub>, y<sub>θ</sub>) are obtained by a translation of the original coordinates  
20 (x<sub>0</sub>, y<sub>0</sub>) and by a rotation of θ,

$$\begin{bmatrix} x_\theta \\ y_\theta \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix}$$

$$G_{\sigma_x, \sigma_y}(x, y) = A \exp \left\{ - \left( \frac{x}{\sqrt{2}\sigma_x} \right)^2 - \left( \frac{y}{\sqrt{2}\sigma_y} \right)^2 \right\}$$

25

A representing the amplitude,  
σ<sub>x</sub> et σ<sub>y</sub>, representing the width of the gaussian envelop along the x  
and y axis respectively.

$$excitation(x, y, \sigma_x, \sigma_y, f, \theta) = \begin{cases} gabor(x, y, \sigma_x, \sigma_y, f, \theta) & \text{if } -1/(4f) \leq x_\theta \leq 1/(4f) \\ 0 & \text{otherwise} \end{cases}$$

In order to obtain elliptic shapes, we take different variances  $\sigma_x < \sigma_y$ .

5 Finally, we get the output:

$$R_{i,j}^{exc}(x, y) = R'_{i,j}(x, y) * excitation(x, y, \sigma_x, \sigma_y, f, \theta)$$

In step S3, the inhibition is calculated by the following formula :

10

$$inhibition(x, y, \sigma_x, \sigma_y, f, \theta) = \begin{cases} 0 & \text{si } -1/(4f) \leq x_\theta \leq 1/(4f) \\ |gabor(x, y, \sigma_x, \sigma_y, f, \theta)| & \text{sinon.} \end{cases}$$

And finally:

15

$$R_{i,j}^{inh}(x, y) = R'_{i,j}(x, y) * inhibition(x, y, \sigma_x, \sigma_y, f, \theta)$$

In step S4, the difference between the excitation and the inhibition is done. The positive components are kept, the negative components are set to "0". This is the following operation,

20

$$R''_{i,j}(x, y) = |R_{i,j}^{exc}(x, y) - R_{i,j}^{inh}(x, y)|_{>0}$$

In step S5, for each orientation, for each sub-band of the second domain, two convolution products are calculated:

25

$$L_{i,j}^0(x, y) = R''_{i,j}(x, y) * B_{i,j}^0(x, y)$$

$$L_{i,j}^1(x, y) = R''_{i,j}(x, y) * B_{i,j}^1(x, y)$$

$B_{i,j}^0(x,y)$  and  $B_{i,j}^1(x,y)$  are 2 half-butterfly filters. The profile of these filters allow the modelling of facilitative interactions for  $\theta=0$  given on figure 9. These filters are defined by using a bipole/butterfly filter.

It consists of a directional term  $D_{\theta}(x,y)$  and a proximity term generated by a circle  $C_r$  blurred by a gaussian filter  $G_{\sigma_x, \sigma_y}(x,y)$ .

$$B_{\theta_{i,j}, \alpha, \gamma, \sigma}(x,y) = D_{\theta_{i,j}}(x,y) \cdot C_r * G_{\sigma_x, \sigma_y}(x,y)$$

$$\text{with } D_{\theta_{i,j}}(x,y) = \begin{cases} \cos(\frac{\pi/2}{\alpha} \varphi) \text{ si } \varphi < \alpha \\ 0 \text{ sinon.} \end{cases}$$

and  $\varphi = \arctan(y'/x')$ , where  $(x', y')^T$  is the vector  $(x, y)^T$  rotated by  $\theta_{i,j}$ . The parameter  $\alpha$  defines the opening angle  $2\alpha$  of the bipole filter. It depends on the angular selectivity  $\gamma$  of the considered sub-band. We take  $\alpha = 0.4 \times \gamma$ . The size of the bipole filter is about twice the size of the CRF of a visual cell.

In step S6, we compute the facilitative coefficient:

$$f_{i,j}^{iso}(x,y) = D\left(\frac{L_{i,j}^1(x,y) + L_{i,j}^0(x,y)}{\max(\beta, |L_{i,j}^1(x,y) - L_{i,j}^0(x,y)|)}\right)$$

with,

$\beta$  a constant,

$$D(z) = \begin{cases} 0, z \leq s_1, \\ \alpha_1, z \leq s_2, \\ \dots \\ \alpha_{N-1}, z \leq s_{N-1} \end{cases} \quad \text{where } \alpha_i \leq 1, i \in [0 \dots N-1]$$

An illustration of the operator  $D(z)$  is given on figure 9.

To ease the application of the facilitative coefficient, the operator  $D(z)$  ensures that the facilitative coefficient is constant by piece as shown on figure 9.



In step S7, the facilitative coefficient is applied to the normalized result obtained in step S4.

$$5 \quad R_{i,j}'''(x,y) = R_{i,j}''(x,y) \times (1 + f_{i,j}^{iso}(x,y))$$

Going back to step E8 of figure 1, after step S7 of figure 7, the four saliency maps obtained for the domain II are combined to get the whole saliency map according to the following formula:

$$10 \quad fixation(x,y) = \alpha \times R_{II,0}'''(x,y) + \beta \times R_{II,1}'''(x,y) + \chi \times R_{II,2}'''(x,y) + \delta \times R_{II,3}'''(x,y)$$

$\alpha, \beta, \chi, \delta$  represent weighting coefficients which depend on the application (watermarking, coding...).

15

In other embodiments, the saliency map can be obtained by a calculation using the whole 17 sub-bands and not only the sub-bands of domain II.

20

Figure 2 represents the general flow-chart of a preferred embodiment of the method according to the invention applied to a colour image.

Steps T1, T4, T'4, T''4, T5 and T8 are done in the spatial domain.

Steps T2, T'2, T''2, T3, T'3, T''3 are done in the frequencial domain.

A Fourier transformation is applied on three components between step

25

T1 and steps T2, T'2, T''2.

An inverse Fourier transformation is applied between respectively T3, T'3, T''3 and T4, T'4 and T''4.

30

Steps T6 and T7 can be done in the frequencial or spatial domain. If they are done in the frequencial domain, a Fourier transformation is done on the signal between steps T5 and T6 and an inverse Fourier transformation is done between steps T7 and T8.

Step T1 consists in converting the RGB luminances into the Krauskopf's opponent-colors space composed by the cardinal directions A, Cr1 and Cr2.

5 This transformation to the opponent-colors space is a way to decorrelate color information. In fact, it's believed that the brain uses 3 different pathways to encode information: the first conveys the luminance signal (A), the second the red and green components (Cr1) and the third the blue and yellow components (Cr2).

10 These cardinal directions are in closely correspondence with signals stemming from the three types of cones (L,M,S).

Each of the three components RGB firstly undergoes a power-law non linearity (called gamma law) of the form  $x^\gamma$  with  $\gamma \approx 2.4$ . This step is necessary in order to take into account the transfer function of the display system. The CIE (French acronym for "commission internationale de l'éclairage") XYZ tristimulus value which form the basis for the conversion to an HVS color space is then computed by the following equation:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0.412 & 0.358 & 0.18 \\ 0.213 & 0.715 & 0.072 \\ 0.019 & 0.119 & 0.95 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$$

20

The response of the (L,M,S) cones are computed as follows :

$$\begin{pmatrix} L \\ M \\ S \end{pmatrix} = \begin{pmatrix} 0.240 & 0.854 & -0.0448 \\ -0.389 & 1.160 & 0.085 \\ -0.001 & 0.002 & 0.573 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

25

From the LMS space, one has to obtain an opponent color space. There is a variety of opponent color space, which differ in the way to combine the different cones responses. From experimental experiments, the color space designed by Krauskopf have been validated and it is given by the following transformation :

$$\begin{pmatrix} A \\ Cr1 \\ Cr2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ -0.5 & -0.5 & 1 \end{pmatrix} \begin{pmatrix} L \\ M \\ S \end{pmatrix}$$

Then, in step T2, a perceptual decomposition is applied to the luminance component. Preliminary to step T2 and further to step T1, the luminance component is transposed into the frequency domain by using known transformations such as the Fourier transformation in order to be able to apply in step T2, the perceptual sub-band decomposition on the image.

The perceptual sub-band decomposition of step T2 is the same as the step E3 of figure 1, and thus will not be described here, as described earlier.

Concerning the decomposition of chromatic components Cr2 and Cr1, of steps T'2 and T''2 as shown on figure 4, the decomposition leads to 5 psychovisual sub-bands for each of these components distributed on 2 crowns. Preliminary to steps T'2, T''2 and further to step T1, the chrominance components are transposed into the frequency domain by using known transformations such as the Fourier transformation in order to be able to apply in step T'2 and T''2, the perceptual sub-band decomposition on the image.

Two domains of spatial frequency are labeled from I to II:

I: spatial frequencies from 0 to 1.5 cycles per degree,

II: spatial frequencies from 1.5 to 5.7 cycles per degree.

In steps T3, T'3 and T''3, a contrast sensitivity function (CSF) is applied.

In step T3, the same contrast sensitivity as in step E4 of figure 1 is performed on the luminance component and thus will not be described here.

In step T'3 and T''3, the same CSF is applied on the two chromatic components Cr1 and Cr2. On the two chromatic components, a two-dimensional anisotropic CSF designed by Le Callet is applied. It is described in document « critères objectifs avec références de qualité visuelle des images couleurs » of Mr Le Callet, university of Nantes, 2001.

This CSF uses two low-pass filters with a cut-off frequency of about 5.5 cycle per degree and 4.1 cycles per degree respectively for Cr1 and Cr2 components.

In order to permit the direct comparison between early visual features stemming from different visual modalities (achromatic and chromatic

components), the sub-bands related to the visibility are weighted. The visibility threshold being defined as the stimulus 's contrast at a particular point for which the stimulus just becomes visible.

5 An inverse Fourier transformation is then applied on the different components (not shown on figure 2) in order to be able to apply the masking in the spatial domain.

10 Then, an intra masking is applied on the different sub-bands for the chromatic components Cr1 and Cr2 during step T'4 and T''4 and for the achromatic component in step T4. This last step has already been explained in the description of figure 1, step E6. Thus, it will not be described again here.

15 Intra channel masking is incorporated as a weighing of the outputs of the CSF function. Masking is a very important phenomenon in perception as it describes interactions between stimuli. In fact, the visibility threshold of a stimulus can be affected by the presence of another stimulus.

20 Masking is strongest between stimuli located in the same perceptual channel or in the same sub-band. We apply the intra masking function designed by Dally on the achromatic component as described on figure 1, step E6 and, on the color component, the intra masking function described in document of P. Le Callet and D. Barba, "Frequency and spatial pooling of visual differences for still image quality assessment", in Proc. SPIE Human Vision and Electronic Imaging Conference, San Jose, CA, Vol. 3959, January 2000.

25 These masking functions consist of non linear transducer as expressed in document of Legge and Foley, "Contrast Masking in Human Vision", Journal of the Optical Society of America, Vol. 70, , N° 12, pp. 1458-1471, December 1980.

30 Visual masking is strongest between stimuli located in the same perceptual channel (intra-channel masking). Nevertheless, as shown in numerous studies, there are several interactions called inter-component masking providing a masking or a pedestal effect. From psychophysics experiments, significant inter-components interactions involving the chromatic components have been elected. Consequently, the sensitivity of achromatic

component could be increased or decreased by the Cr1 component. The influence of the Cr2 on the achromatic component is considered insignificant. Finally, the Cr1 can also modulate the sensitivity of Cr2 component (and vice versa).

5

Then in step T5, a chromatic reinforcement is done.

10 The colour is one of the strongest attractor of the attention and the invention wants to take advantage of this attraction strength by putting forward the following property: the existence of regions showing a sharp colour and fully surrounded of areas having quite other colours implies a particular attraction to the borders of this region.

15 To avoid the difficult issue of aggregating measures stemming from achromatic and chromatic components, the colour facilitation consists in enhancing the saliency of achromatic structure by using a facilitative coefficient computed on the low frequencies of the chromatic components.

20 In the preferred embodiment, only a sub-set of the set of achromatic channels is reinforced. This subset contains 4 channels having an angular selectivity equal to  $\pi/4$  and a spatial radial frequency (expressed in cyc/deg) belonging to  $[1.5, 5.7]$ . One notes these channels  $R_{i,j}$  where  $i$  represents the spatial radial frequency and  $j$  pertains to the orientation. In our example,  $j$  is equal to  $\{0, \pi/4, \pi/2, 3\pi/4\}$ . In order to compute a facilitative coefficient, one determines for each pixel of the low frequency of

25 Cr1 and Cr2 the contrast value related to the content of adjacent areas and to the current orientation of the reinforced achromatic channel as illustrated on the figure 11. On figure 11, the contrast value is obtained by computing the absolute difference between the average value of the set A and the average value of the set B. The sets A and B belong to the low frequency of Cr1 or Cr2

30 and are oriented in the preferred orientation of the considered achromatic channel.



The chromatic reinforcement is achieved by the following equation, on an achromatic (luminance) channel  $R_{i,j}(x,y)$ .

$$R_{i,j}^{(1)}(x,y) = R_{i,j}(x,y) \times (1 + |A - B|_{Cr1} + |A - B|_{Cr2}) \Big|_{i=H}$$

5

where,

$R_{i,j}^{(1)}(x,y)$  represents the reinforced achromatic sub-band,

$R_{i,j}(x,y)$  represents an achromatic sub-band.

$|A - B|_k$  represents the contrast value computed around the current

10 point on the chromatic component  $k$  in the preferred orientation of the sub-band  $R_{i,j}(x,y)$ , as shown on figure 7. In the embodiment, the sets A and B belong to the sub-band of the first crown (low frequency sub-band) of the chromatic component  $k$  with an orientation equal to  $\pi/4$ .

15

In other embodiments, all the sub-bands can be considered.

In step T6, a center/surround suppressive interaction is carried on. This operation consists first in a step of inhibition/excitation.

20

A two-dimensional difference-of-Gaussians (DoG) is used to model the non-CRF inhibition behavior of cells. The  $DoG_{\sigma_x^{ex}, \sigma_y^{ex}, \sigma_x^{inh}, \sigma_y^{inh}}(x,y)$  is given by the following equation :

$$DoG_{\sigma_x^{ex}, \sigma_y^{ex}, \sigma_x^{inh}, \sigma_y^{inh}}(x,y) = G_{\sigma_x^{inh}, \sigma_y^{inh}}(x,y) - G_{\sigma_x^{ex}, \sigma_y^{ex}}(x,y)$$

$$\text{with } G_{\sigma_x, \sigma_y}(x,y) = \frac{1}{2\pi(\sigma_x\sigma_y)^2} \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}\right) \text{ a two-dimensional}$$

25

gaussian.

Parameters  $(\sigma_x^{ex}, \sigma_y^{ex})$  and  $(\sigma_x^{inh}, \sigma_y^{inh})$  correspond to spatial extends of the Gaussian envelope along the  $x$  and  $y$  axis of the central Gaussian (the CRF center) and of the inhibitory Gaussian (the surround) respectively. These parameters have been experimentally determined in accordance with the

radial frequency of the second crown (the radial frequency  $f \in [1.5, 5.7]$  is expressed in cycles/degree). Finally, the non-classical surround inhibition can be modeled by the normalized weighting function  $w_{\sigma_x^{ex}, \sigma_y^{ex}, \sigma_x^{inh}, \sigma_y^{inh}}(x, y)$  given by the following equation:

$$5 \quad w_{\sigma_x^{ex}, \sigma_y^{ex}, \sigma_x^{inh}, \sigma_y^{inh}}(x, y) = \frac{1}{\|H(DoG_{\sigma_x^{ex}, \sigma_y^{ex}, \sigma_x^{inh}, \sigma_y^{inh}})\|_1} H(DoG_{\sigma_x^{ex}, \sigma_y^{ex}, \sigma_x^{inh}, \sigma_y^{inh}}(x', y'))$$

with,

$$H(z) = \begin{cases} 0, & z < 0 \\ z, & z \geq 0 \end{cases}$$

$(x', y')$  is obtained by translating the original coordinate system by  $(x_0, y_0)$  and rotating it by  $\theta_{i,j}$  expressed in radian,

$$10 \quad \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos \theta_{i,j} & \sin \theta_{i,j} \\ -\sin \theta_{i,j} & \cos \theta_{i,j} \end{bmatrix} \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix},$$

$\|\cdot\|_1$  denotes the  $L_1$  norm, i.e the absolute value.

The figure 12 shows the structure of non-CRF inhibition.

15 The figure 13 shows a profile example of the normalized weighting function  $w_{\sigma_x^{ex}, \sigma_y^{ex}, \sigma_x^{inh}, \sigma_y^{inh}}(x, y)$ .

The response  $R_{i,j}^{(2)}(x, y)$  of cortical cells to a particular sub-band  $R_{i,j}^{(1)}(x, y)$  is computed by the convolution of the sub-band  $R_{i,j}^{(1)}(x, y)$  with the weighting function  $w_{\sigma_x^{ex}, \sigma_y^{ex}, \sigma_x^{inh}, \sigma_y^{inh}}(x, y)$ :

$$20 \quad R_{i,j}^{(2)}(x, y) = H(R_{i,j}^{(1)}(x, y) - R_{i,j}^{(1)}(x, y) * w_{\sigma_x^{ex}, \sigma_y^{ex}, \sigma_x^{inh}, \sigma_y^{inh}}(x, y))|_{i=11}$$

with  $H(z)$  defined as has been described above.

In step T7, a facilitative interaction is carried on.

25 This facilitative interaction is usually termed contour enhancement or contour integration.

Facilitative interactions appear outside the CRF along the preferred orientation axis. These kinds of interactions are maximal when center and

surround stimuli are iso-oriented and co-aligned. In other words, as shown by several physiologically observations, the activity of cell is enhanced when the stimuli within the CRF and a stimuli within the surround are linked to form a contour.

5 Contour integration in early visual preprocessing is simulated using two half butterfly filter  $B_{i,j}^0$  and  $B_{i,j}^1$ . The profiles of these filters are shown on the 9 and they are defined by using a bipole/butterfly filter. It consists of a directional term  $D_\theta(x,y)$  and a proximity term generated by a circle  $C_r$  blurred by a gaussian filter  $G_{\sigma_x, \sigma_y}(x,y)$ .

10

$$B_{\theta_{i,j}, \alpha, r, \sigma}(x, y) = D_{\theta_{i,j}}(x, y) \cdot C_r * G_{\sigma_x, \sigma_y}(x, y)$$

$$\text{with } D_{\theta_{i,j}}(x, y) = \begin{cases} \cos(\frac{\pi/2}{\alpha} \varphi) \text{ si } \varphi < \alpha \\ 0 \text{ sinon.} \end{cases}$$

and  $\varphi = \arctan(y'/x')$ , where  $(x', y')^T$  is the vector  $(x, y)^T$  rotated by  $\theta_{i,j}$ . The parameter  $\alpha$  defines the opening angle  $2\alpha$  of the bipole filter. It depends on the angular selectivity  $\gamma$  of the considered sub-band. One takes  $\alpha = 0.4 \times \gamma$ . The size of the bipole filter is about twice the size of the CRF of a visual cell.

15

The two half butterfly filter  $B_{i,j}^0$  and  $B_{i,j}^1$  are after deduced from the butterfly filter by using appropriate windows.

20

For each orientation, sub-band and location, one computes the facilitative coefficient :

$$f_{i,j}^{iso}(x, y) = D \left( \frac{L_{i,j}^1(x, y) + L_{i,j}^0(x, y)}{\max(\beta, |L_{i,j}^1(x, y) - L_{i,j}^0(x, y)|)} \right)$$

with,

25

$\beta$  a constant,

$$L_{i,j}^0(x, y) = R_{i,j}^{(2)}(x, y) * B_{i,j}^0(x, y),$$

$$L_{i,j}^1(x, y) = R_{i,j}^{(2)}(x, y) * B_{i,j}^1(x, y),$$



$$D(z) = \begin{cases} 0, z \leq s_1, \\ \alpha_1, z \leq s_2, \\ \dots \\ \alpha_{N-1}, z \leq s_{N-1} \end{cases} \quad \text{where}$$

An illustration of the operator  $D(z)$  is given on figure 9.

5 The sub-band  $R_{i,j}^{(3)}$  resulting from the facilitative interaction is finally obtained by weighting the sub-band  $R_{i,j}^{(2)}$  by a factor depending on the ratio of the local maximum of the facilitative coefficient  $f_{i,j}^{iso}(x,y)$  and the global maximum of the facilitative coefficient computed on all sub-bands belonging to the same range of spatial frequency :

10

$$R_{i,j}^{(3)}(x,y) = R_{i,j}^{(2)}(x,y) \times (1 + \eta^{iso} \times \frac{\max_{(x,y)}(f_{i,j}^{iso}(x,y))}{\max_j(\max_{(x,y)}(f_{i,j}^{iso}(x,y)))}) \Big|_{i=II}$$

From a standard butterfly shape, this facilitative factor permits to improve the saliency of isolated straight lines.  $\eta^{iso}$  permits to control the strength of this facilitative interaction.

15

In step E8, the saliency map is obtained by summing all the resulting sub-bands obtained in step E7.

$$S(x,y) = \sum_{i=II,j} R_{i,j}^{(3)}(x,y)$$

20

In other embodiments of the invention, one can use all the sub-bands and not only the sub-bands of the second crown.

25

Although cortical cells tuned to horizontal and vertical orientations are almost as numerous as cells tuned to other orientations, we don't introduce any weighting. This feature of the HVS is implicitly mimic by the application of 2D anisotropic CSF.

## Claims

1. Method for creating a saliency map of an image characterized in that it  
5 comprises the steps of :

- Projection (E1) of said image according to the luminance (A) component and if said image is a color image, according to the luminance (A) component and according to the chrominance components (Cr1, Cr2),
- 10 - Perceptual sub-bands decomposition (E3, T2, T'2, T''2) of said components (A, Cr1, Cr2) according to the visibility threshold of a human eye,
- Extraction (E7) of the salient elements of the sub-bands related to the luminance (A) component,
- 15 - Contour enhancement (E8, T7) of said salient elements in each sub-band related to the luminance (A) component,
- Calculation (T7) of a saliency map from the contour enhancement, for each sub-band related to the luminance (A) component.
- 20 - Creation (T8) of the saliency map as a function of the saliency maps obtained for each sub-band.

2. Method according to claim 1 characterized in that it comprises, further  
to the perceptual sub-bands decomposition,

- 25 - a step of achromatic contrast sensitivity function (CSF) for the luminance (A) component and if said image is a color image,
- a step of chromatic sensitivity function for the chromatic components (Cr1, Cr2).

30 3. Method according to claim 2 characterized in that it comprises a step (E6, T4, T'4, T''4) of visual masking, further to the step of contrast sensitivity function, for each sub-band of the luminance (A) component and of the chromatic (Cr1, Cr2) components.

4. Method according to claim 3 characterized in that, when said image is a color image, it comprises a step (T5) of chromatic reinforcement of the luminance (A) sub-bands.

5. Method according to any of the preceding claims characterized in that the perceptual sub-bands decomposition is obtained by carving-up the frequency domain both in spatial radial frequency and orientation.

6. Method according to claim 5 characterized in that the perceptual decomposition of the luminance (A) component leads to 17 psycho visual sub-bands distributed on four crowns.

7. Method according to claim 5 or 6 characterized in that the perceptual decomposition of the chromatic components (Cr1, Cr2) leads to 5 psycho visual sub-bands distributed on two crowns for each chromatic component (Cr1, Cr2).

8. Method according to claims 4 to 7 characterized in that the chromatic reinforcement of the luminance (A) component is done on the sub-bands of the second crown and based on the sub-bands of the first crown of the chromatic components (Cr1, Cr2).

9. Device for creating a saliency map of an image characterized in that it comprises means for:

- 25 - Projecting said image according to the luminance (A) component and if said image is a color image, according to the luminance (A) component and according to the chrominance components (Cr1, Cr2),
- 30 - Transposing into the frequential domains said luminance and chrominance signals,
- Decomposing into perceptual sub-bands said components of the frequential domain according to the visibility threshold of a human eye,

23

5

- Extracting the salient elements of the sub-bands related to the luminance component,
- Contour enhancing said salient elements in each sub-band related to the luminance component,
- Calculating a saliency map from the contour enhancement, for each sub-band related to the luminance component.
- Creating the saliency map as a function of the saliency maps obtained for each sub-band.

10

# Device and method for creating a saliency map of an image.

## Abstract

5 The invention concerns a device and a method for creating a saliency map of an image. It comprises the steps of :

- 10 - Projection (E1) of said image according to the luminance (A) component and if said image is a color image, according to the luminance (A) component and according to the chrominance components (Cr1, Cr2),
- Perceptual sub-bands decomposition (E3, T2, T'2, T''2) of said components (A, Cr1, Cr2) according to the visibility threshold of a human eye,
- 15 - Extraction (E7) of the salient elements of the sub-bands related to the luminance (A) component,
- Contour enhancement (E8, T7) of said salient elements in each sub-band related to the luminance (A) component,
- Calculation (T7) of a saliency map from the contour enhancement, for each sub-band related to the luminance (A) component.
- 20 - Creation (T8) of the saliency map as a function of the saliency maps obtained for each sub-band.

Fig 2

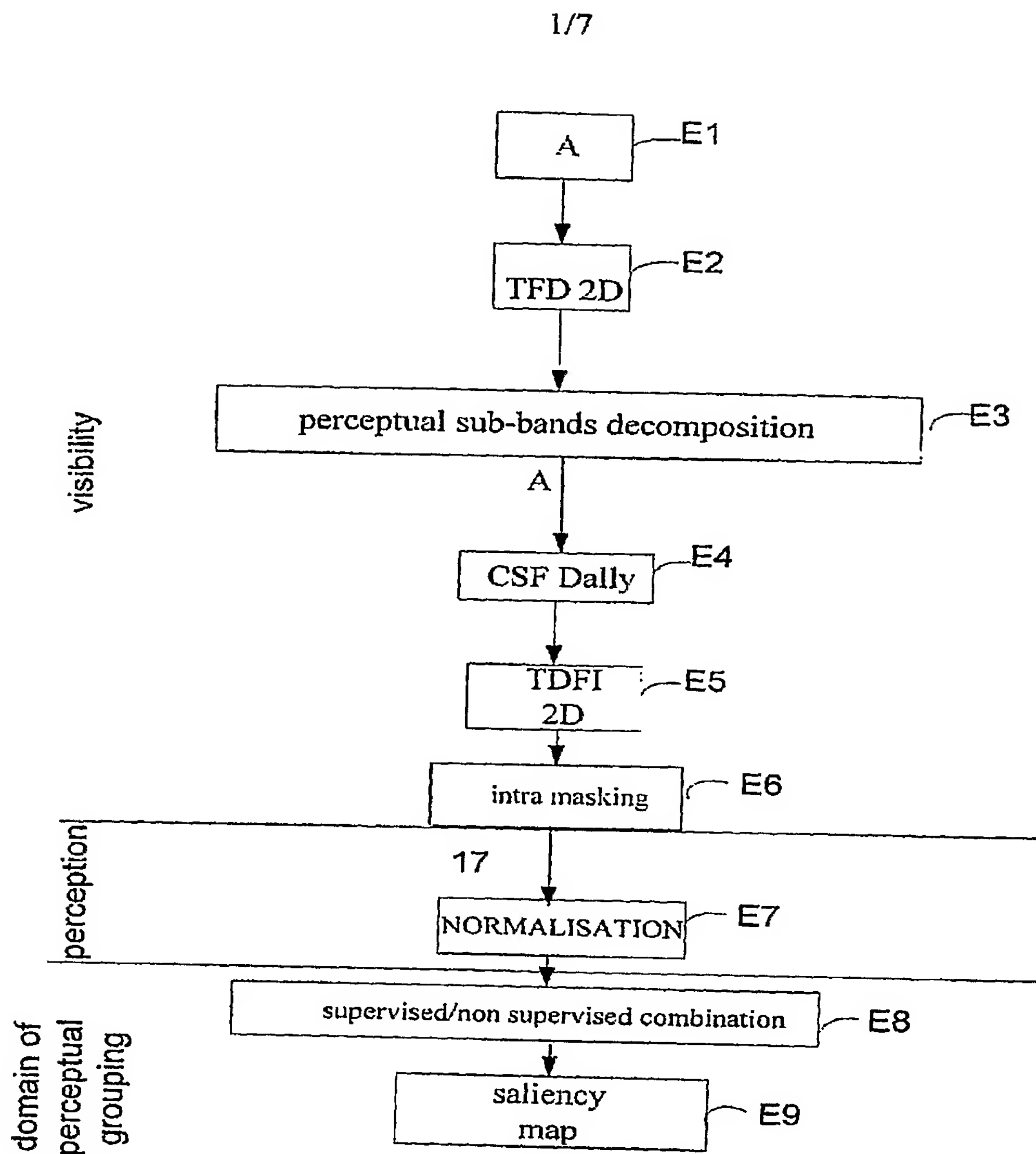


Fig 1

2/7

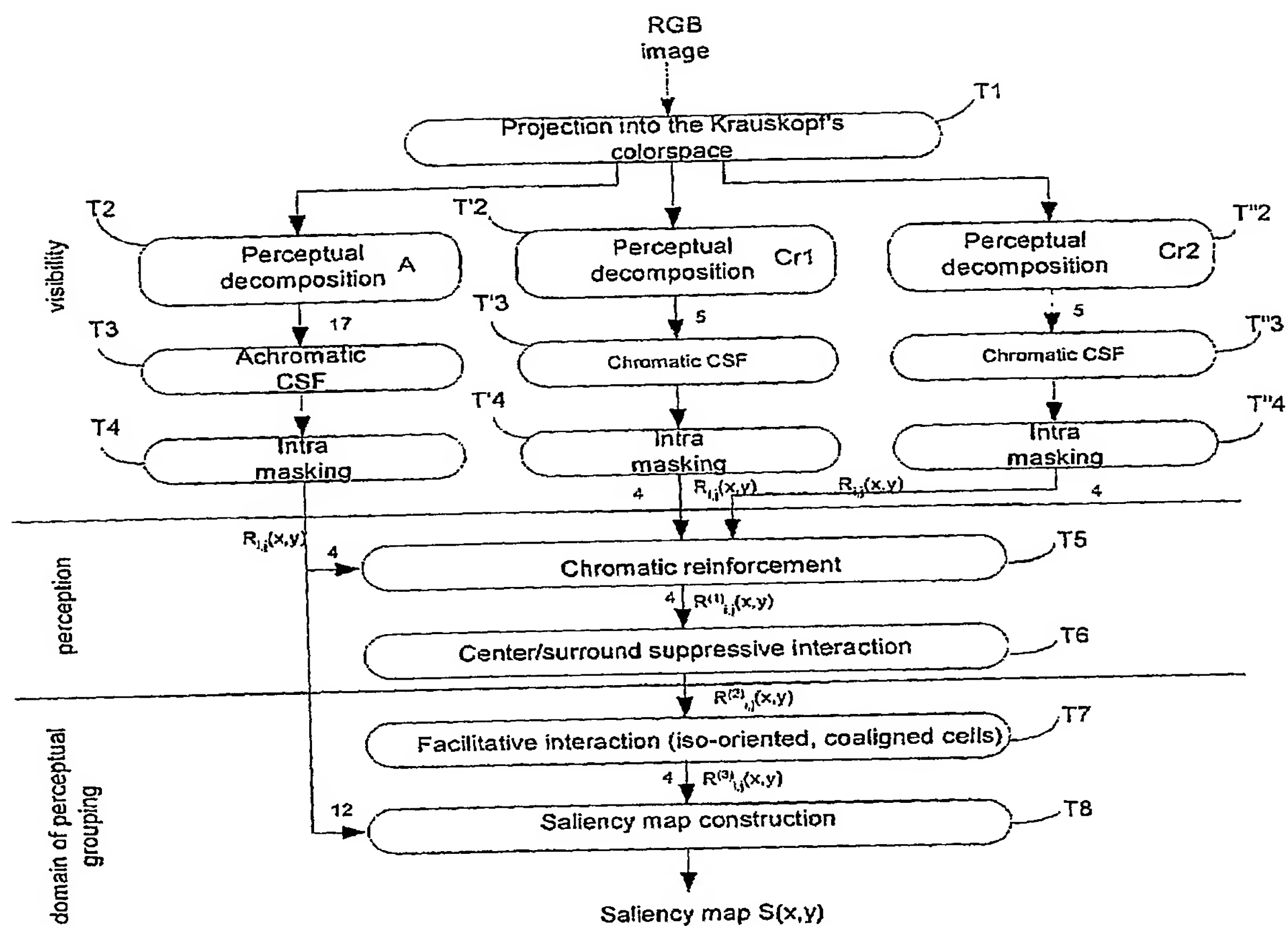


Fig 2

3/7

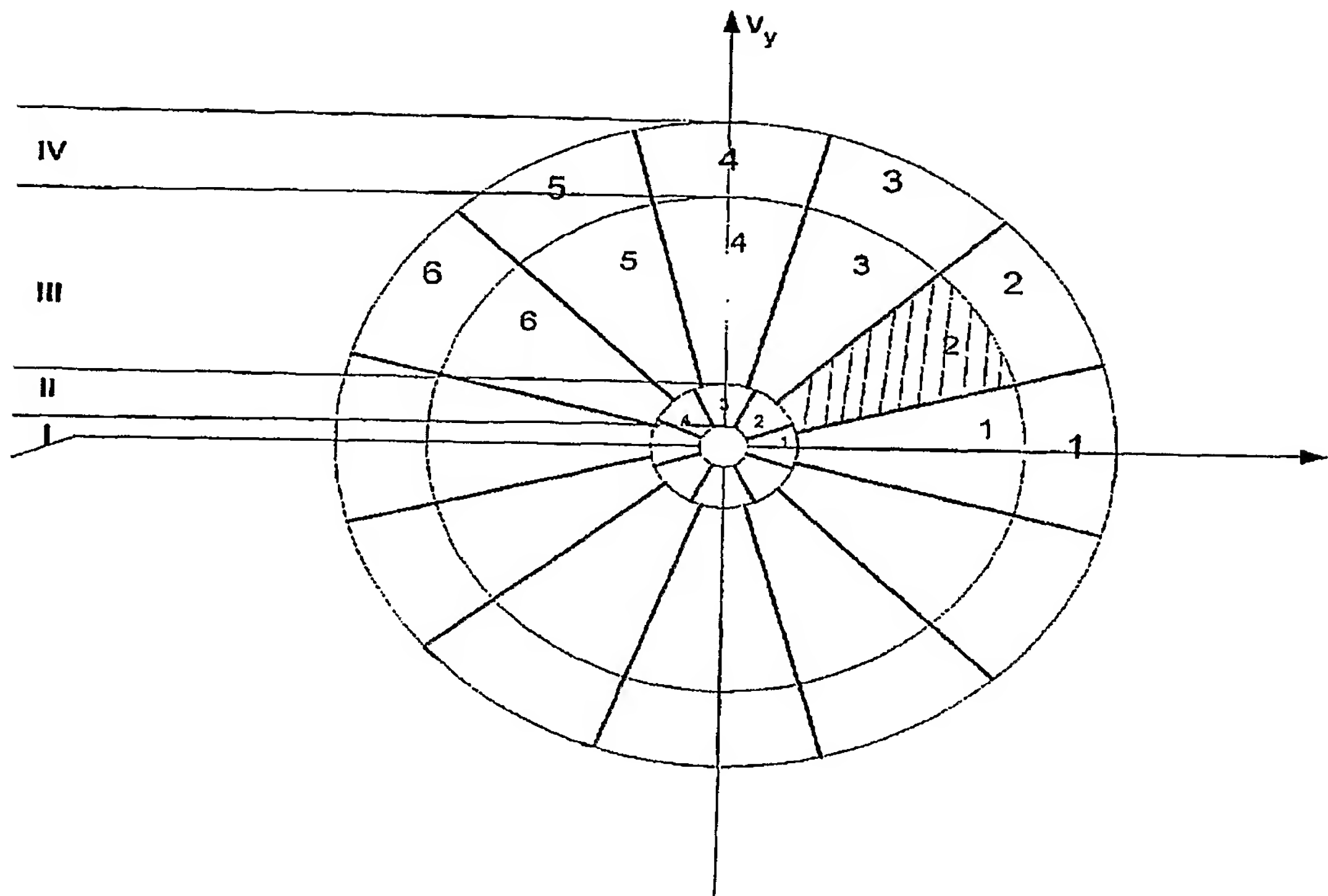


Fig 3

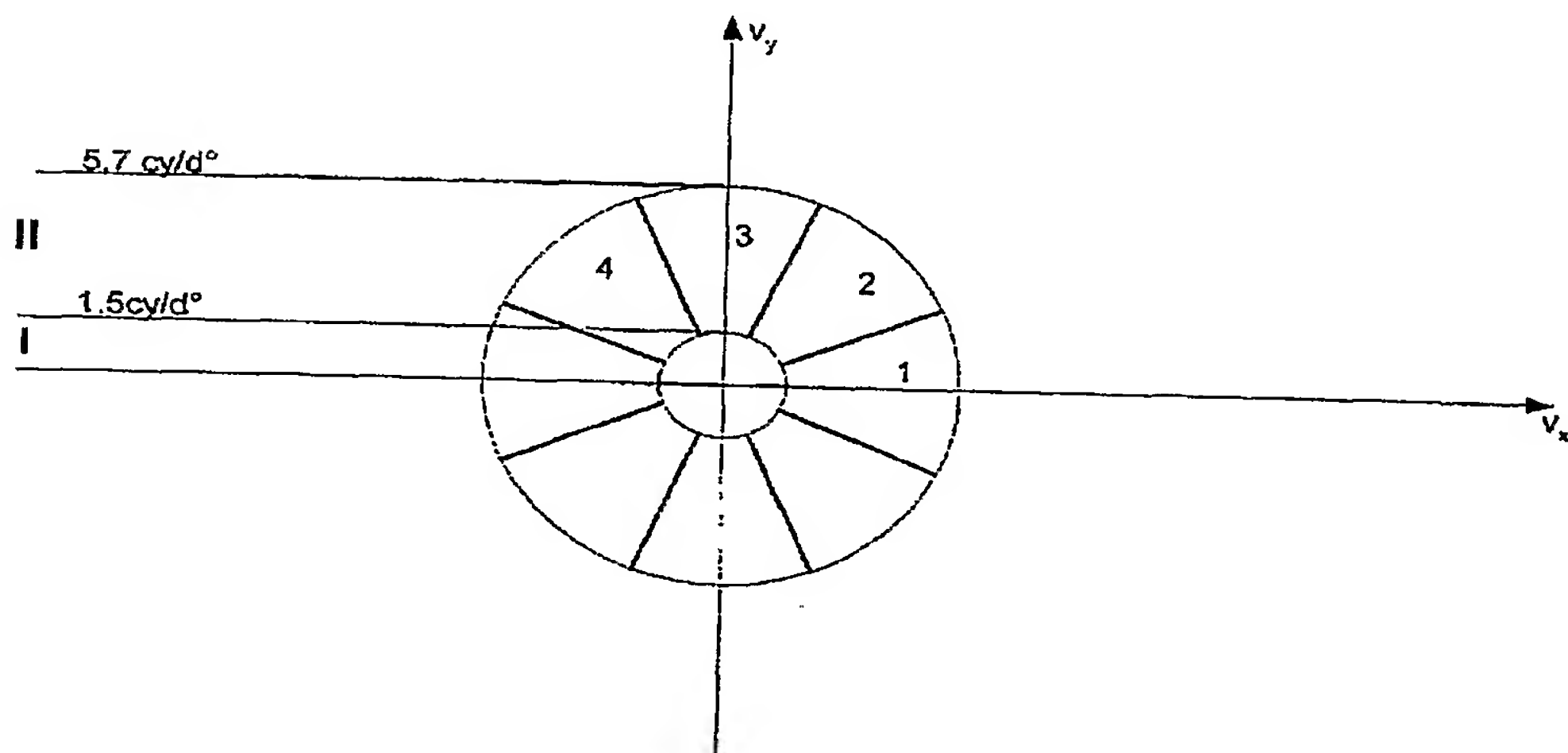


Fig 4



4/7

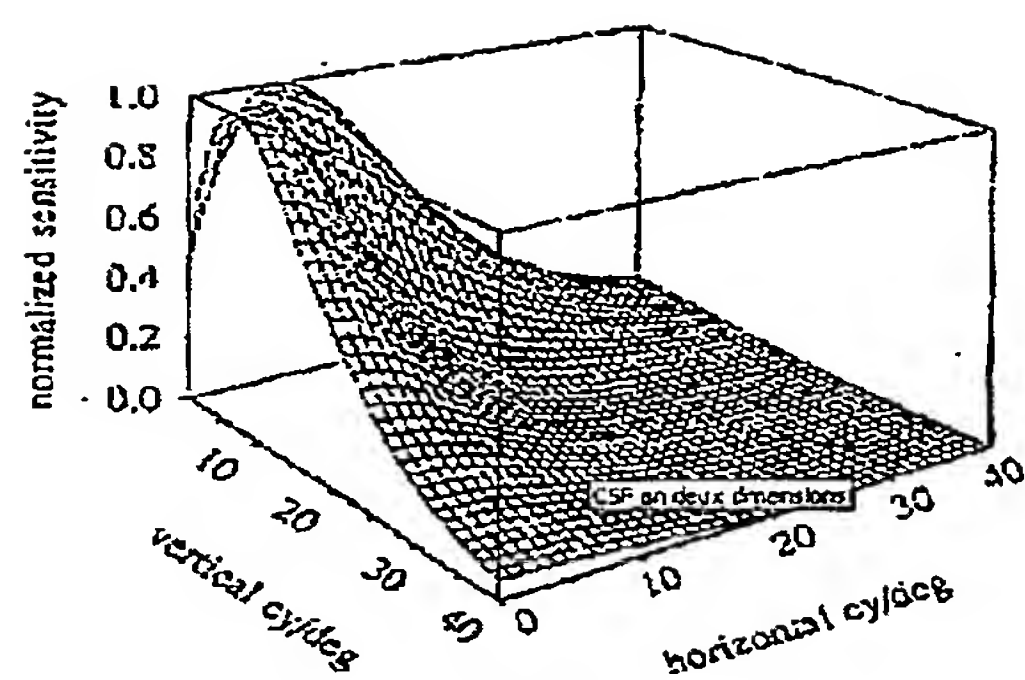


Fig 5

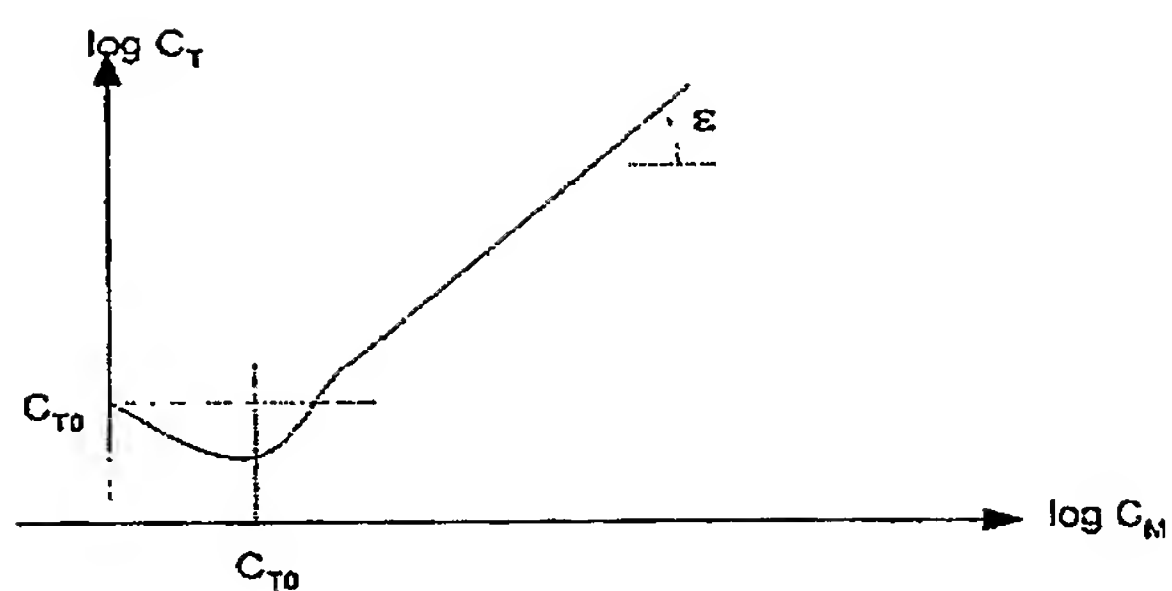


Fig 6a

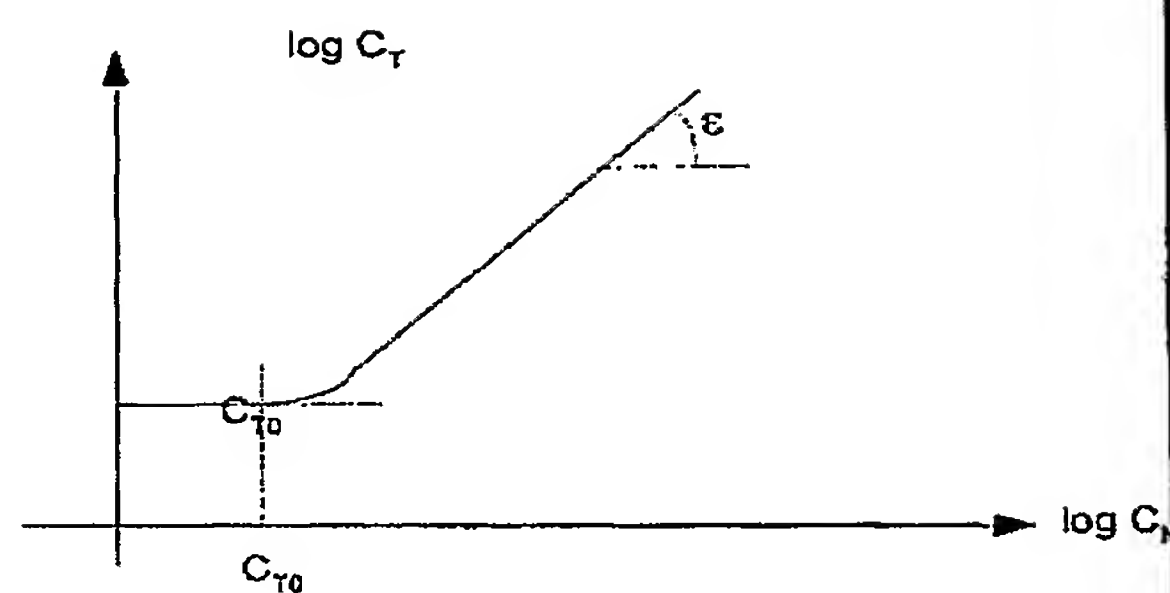


Fig 6b

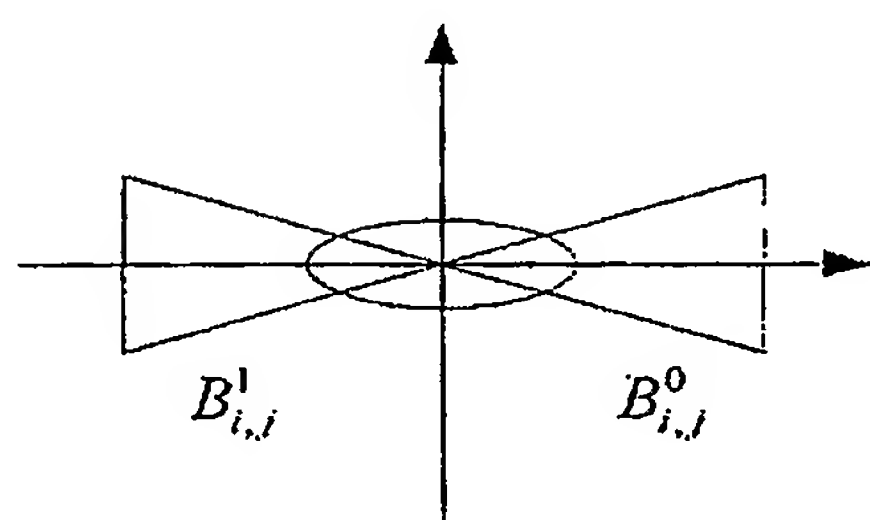


Fig 9

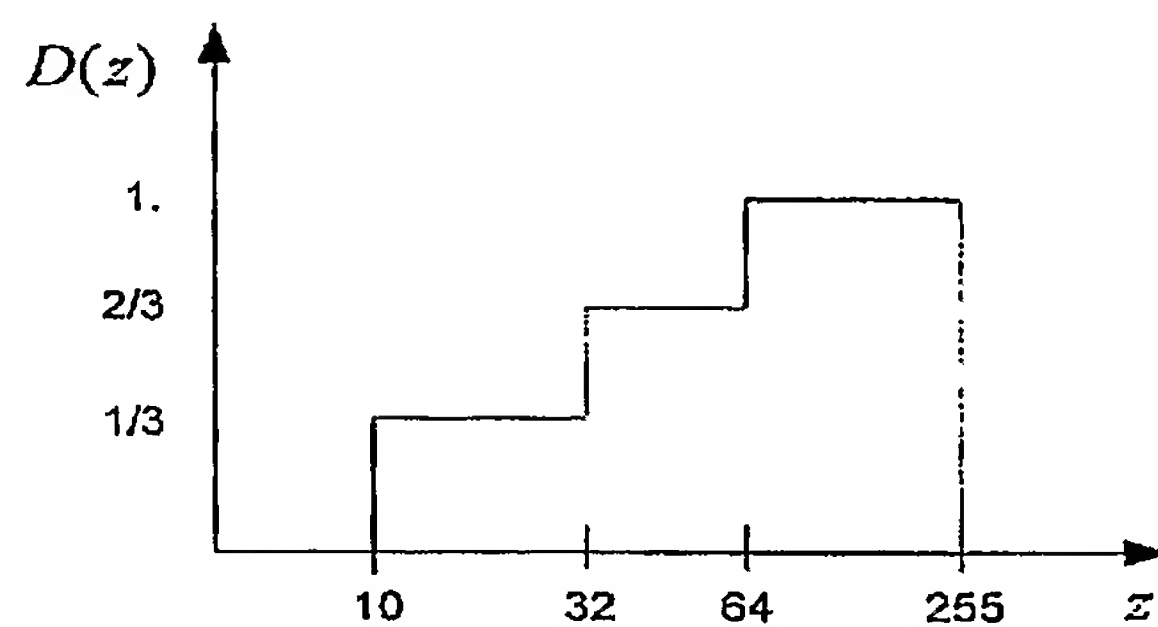


Fig 10

5/7

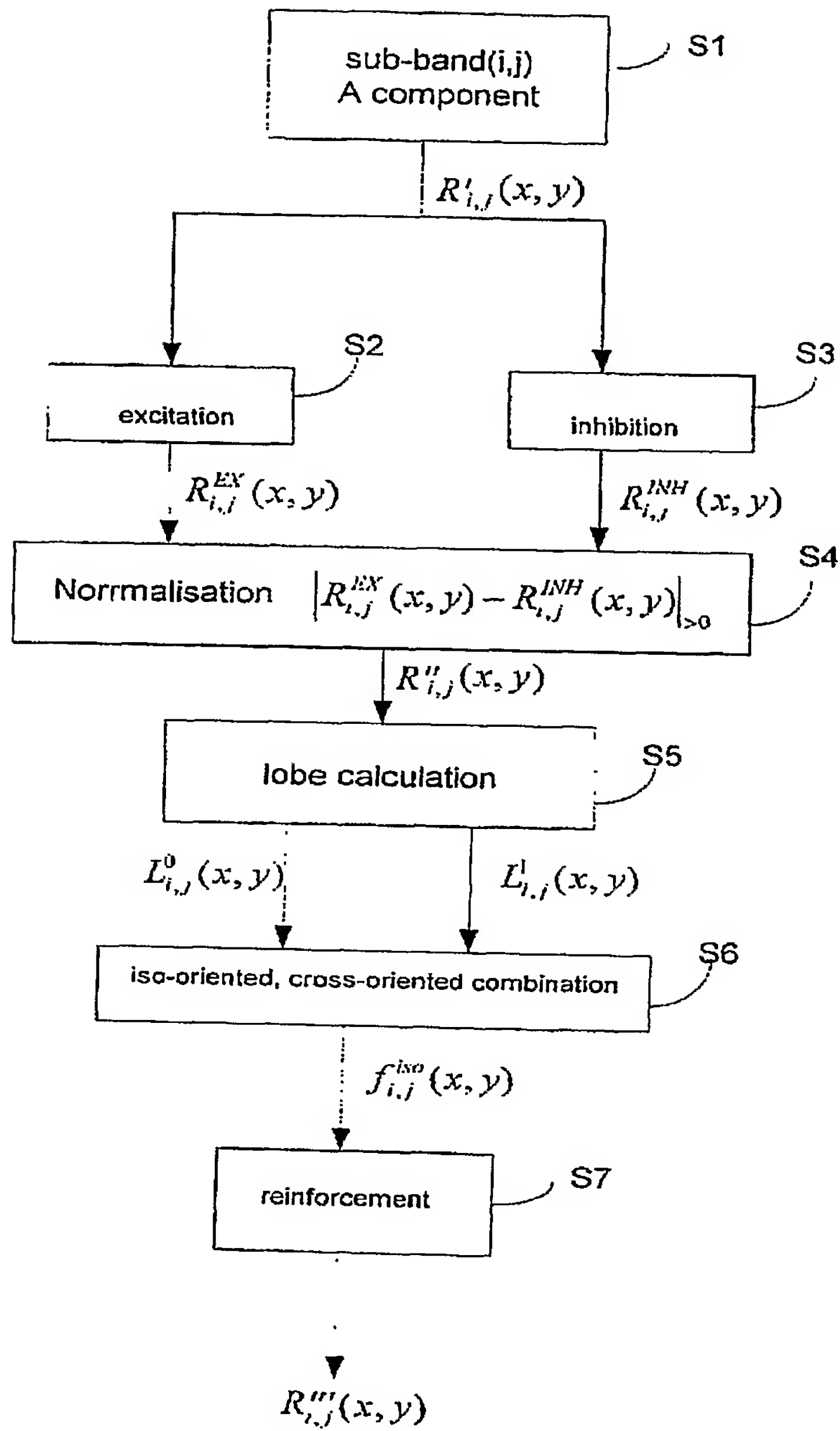


Fig 7

6/7

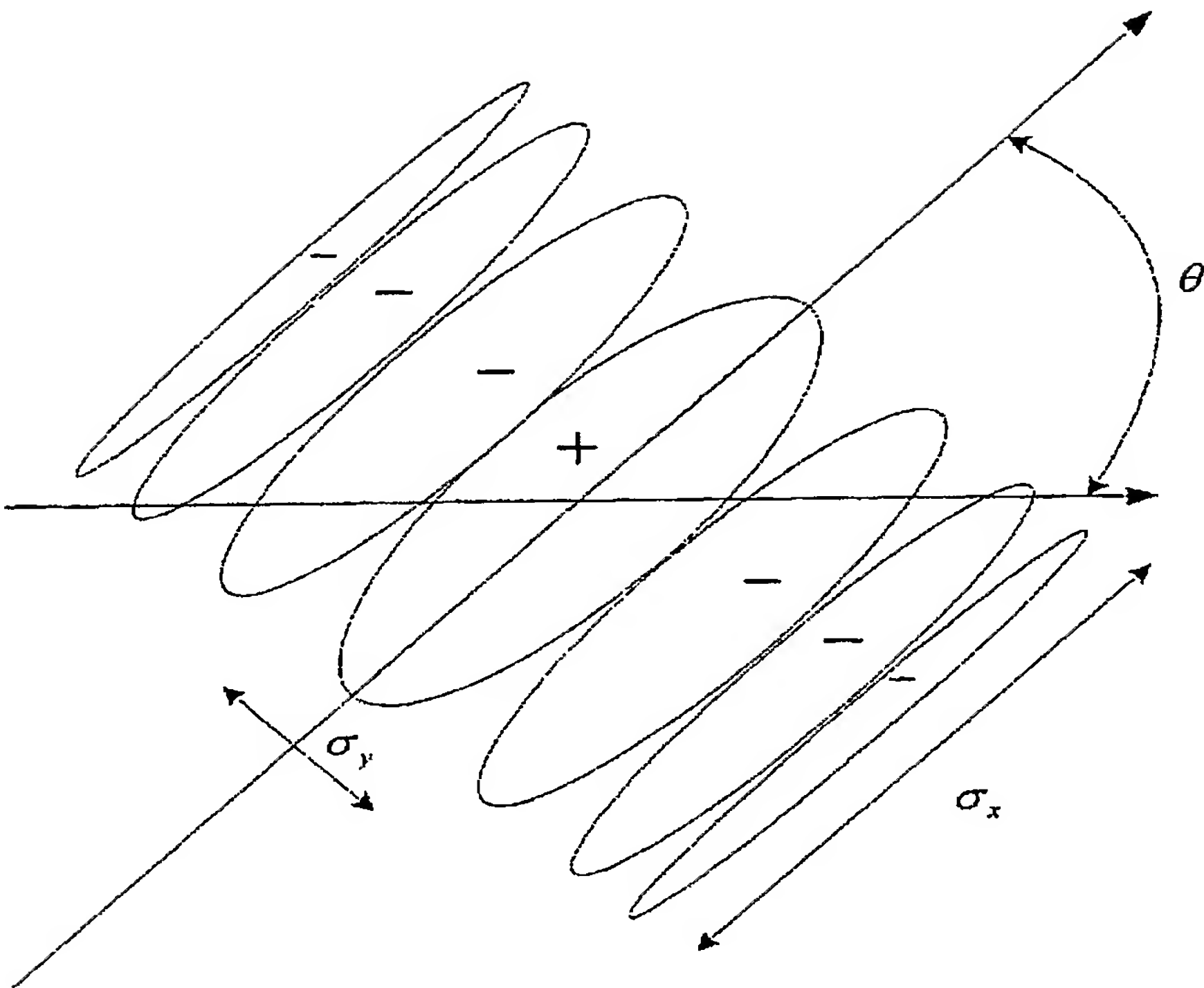


Fig 8

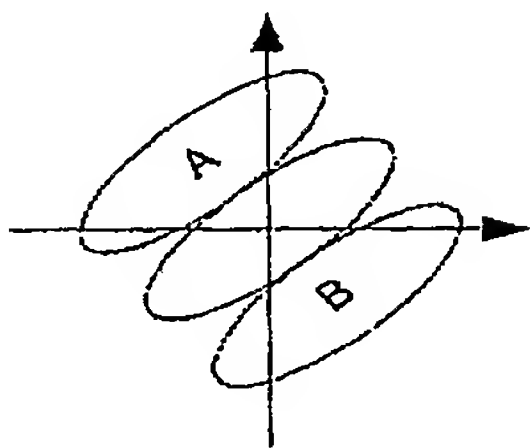


Fig 11

7/7

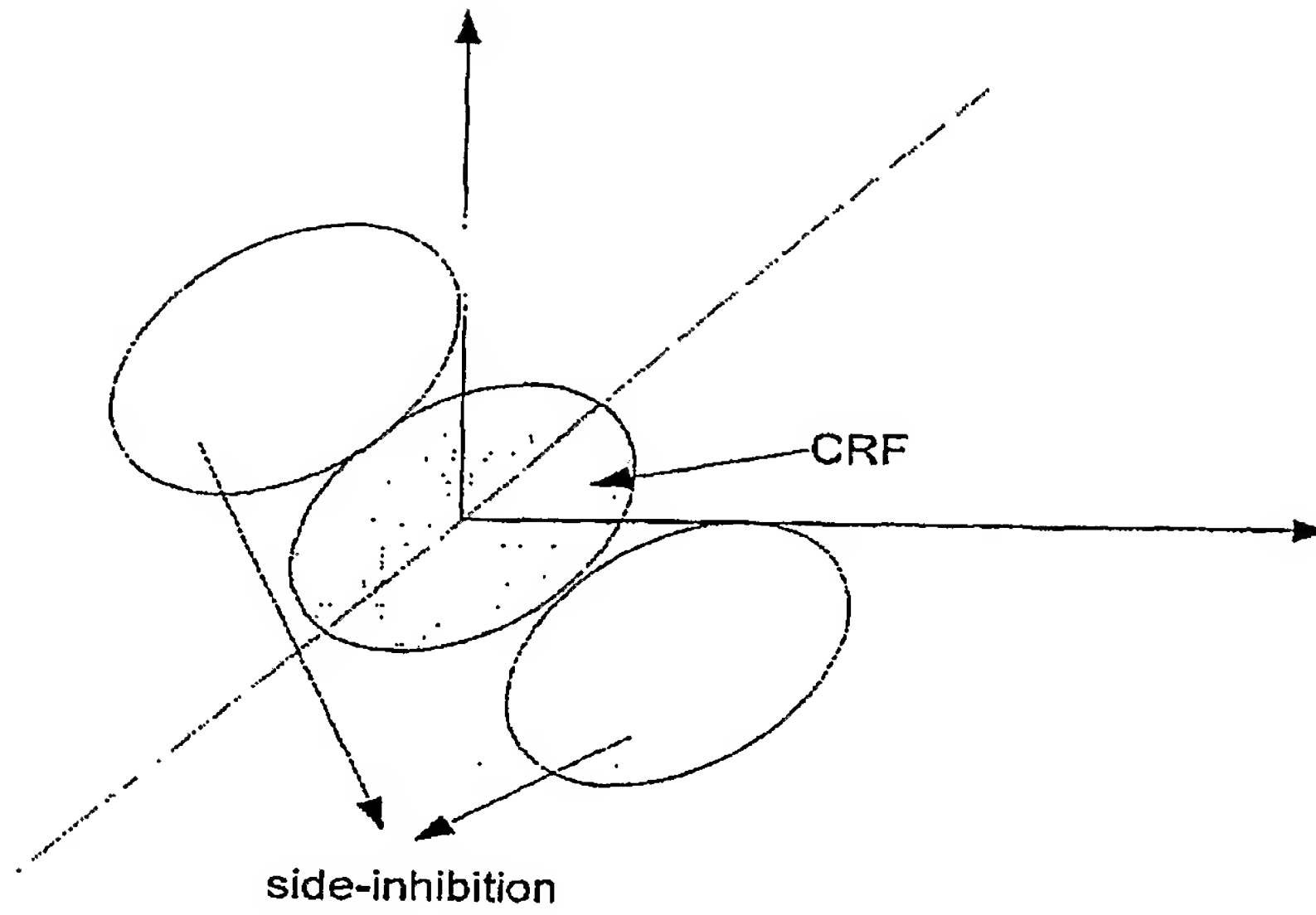


Fig 12

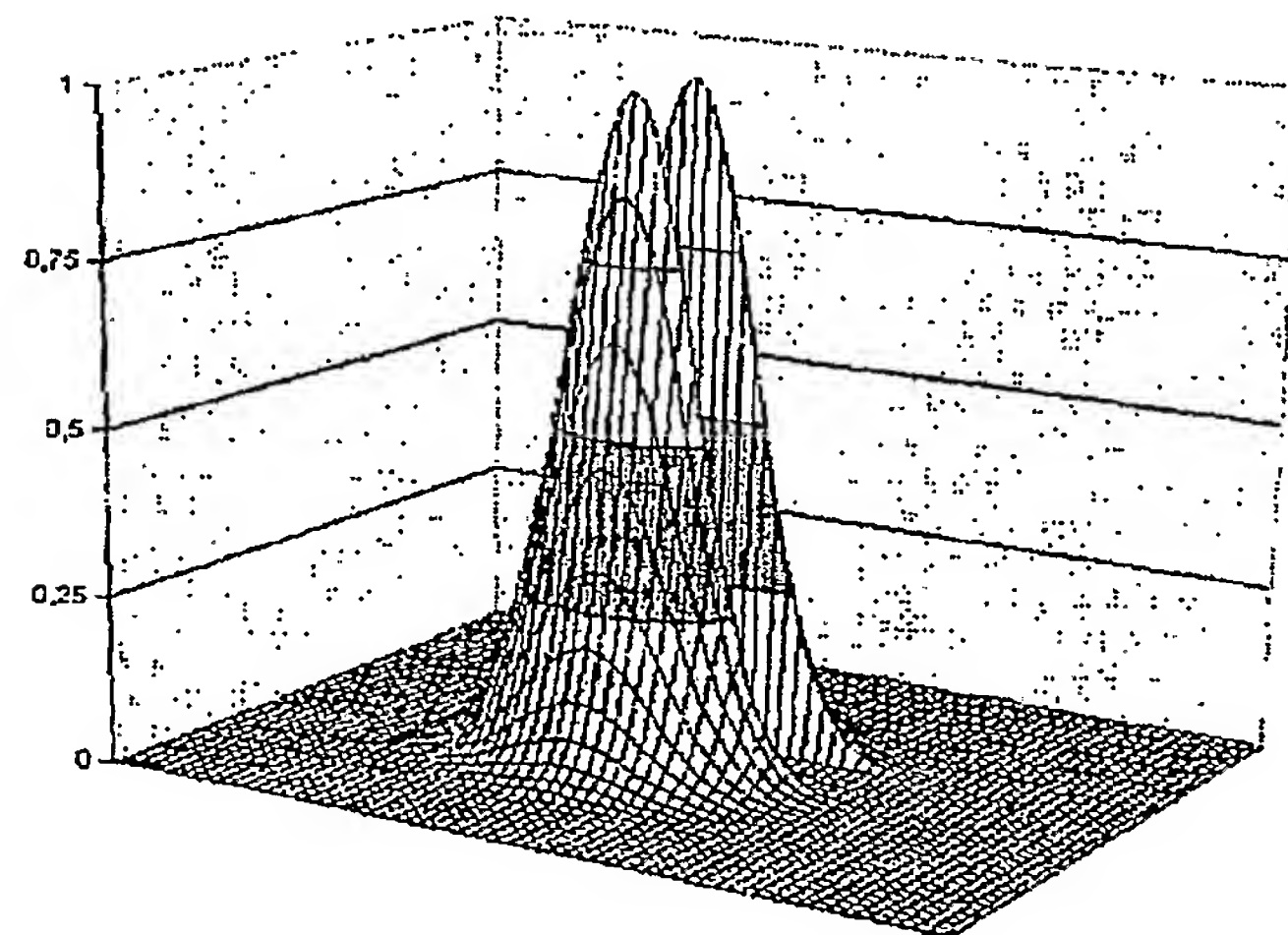


Fig 13

